REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) April 2000	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) June 1999-February 2000
4. TITLE AND SUBTITLE Direct Observation in the Conduct of Training Impact Analyses		5a. CONTRACT OR GRANT NUMBER
		5b. PROGRAM ELEMENT NUMBER 62785
6. AUTHORS Kenneth L. Evans and Jean L. Dyer (U.S. Army Research Institute)		5c. PROJECT NUMBER A790
		5d. TASK NUMBER 204
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences Infantry Forces Research Unit P.O. Box 52086 Fort Benning, GA 31995-2086		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Research Report 1757

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

Important training issues are sometimes not considered when examining the relative merits of competing candidates for an operational system requirement. This is particularly true early in the product development cycle. To address this concern, training impact analysis methods were developed and implemented within the context of an Operational Test (OT) of antitank weapon systems and an Advanced Concept Technology Demonstration (ACTD) of off-the-shelf technologies for urban operations. Data collected were predominately observational, consisting of time-referenced specimen records documented sequentially within their naturally occurring context. These data were used to identify and compare the tasks soldiers had to learn and perform with different candidate systems. Subjective judgments were made about the relative complexity and difficulty of tasks across systems. Relative to a baseline technology or predecessor system, each candidate was ultimately judged to have either a positive, neutral, or negative potential impact on the institutional and unit training base. Training impact rankings of systems were based on the relative numbers of tasks involved, the relative complexity and difficulty of each task, and the relative levels of training resources needed to achieve operational proficiency. Finally, several potential uses of training impact information are suggested.

15. SUBJECT TERMS

Training   Test and Evaluation   Research Methods   Field Observation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	44	Dr. Ken Evans (706) 545-2565

U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1757

Direct Observation in the Conduct of Training Impact Analyses

Kenneth L. Evans and Jean L. Dyer
U.S. Army Research Institute

April 2000

Research Report 1757

Direct Observation in the Conduct of Training Impact Analyses

Kenneth L. Evans and Jean L. Dyer
U.S. Army Research Institute

Infantry Forces Research Unit
Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

April 2000

Army Project Number
20262785A790

Personnel Performance and Training Technology

FOREWORD

Source selection decisions typically do not give training issues sufficient weight when the decisions must be made relatively early in the product development cycle. In part, this situation may be attributable to the lack of a timely and appropriate methodology for determining the potential impact of a candidate system on the institutional and unit training base. The training impact analysis guidelines presented in this report directly address that methodological shortcoming.

Our development of training impact analysis methods began over a decade ago, when we examined important training issues related to an Operational Test (OT) of antitank weapon system candidates. More recently, these methods were extended and refined as part of an Advanced Concept Technology Demonstration (ACTD) of off-the-shelf technologies for Military Operations in Urban Terrain (MOUT). An outgrowth of these two analytical projects, this report provides a description of the procedures we used in those projects, as well as the rationale behind them. Without referring to potentially sensitive evaluation information, we offer illustrative examples of our analyses and findings for the benefit of other researchers, training developers, and decision makers involved in the source selection process.

The series of training impact analyses performed in conjunction with the OT and ACTD were documented in separate reports and briefings to their respective sponsors. In the case of the antitank weapons OT, formal briefings were presented to the Infantry School in August 1988 and to the Program Manager-Advanced Antitank Weapon Systems in October of 1988. Over the course of ten MOUT ACTD experiments, a series of separate training impact analysis reports and briefings were given to the Infantry School's Dismounted Battlespace Battle Lab, the Marine Corps Warfighting Lab, and the MOUT ACTD Program Office between February 1998 and June 1999.

Training impact methods can be adapted to a variety of research situations, as the reader shall see. Further, training impact analysis results can be used for purposes other than making more informed source selection decisions. For example, training impact information can give training developers a head start in the design of training programs, devices, and materials prior to the acquisition and fielding of new systems. This kind of information can also help to identify deficiencies in system design, which can sometimes be corrected if identified early in product development.

Zita M. Simutis

ZITA M. SIMUTIS
Technical Director

DIRECT OBSERVATION IN THE CONDUCT OF TRAINING IMPACT ANALYSES

EXECUTIVE SUMMARY

Research Requirement:

Military test and evaluation programs sometimes fail to consider important training issues when examining the relative merits of competing candidates for a particular operational system requirement. This is particularly true early in the product development cycle. For this reason, systems selected for procurement tend to be chosen on the basis of other factors such as military utility, effectiveness, mobility, soldier acceptance, safety, and cost. Yet, competing candidates may be clearly different in the likely impacts each would have on the institutional and unit training base if selected for acquisition and fielding. In most cases, these differential training impacts, whether positive or negative, can be estimated quite early in development. Those required to make source selection decisions need to have timely and accurate training impact information in order to best gauge the merits of competing systems.

Procedure:

Methods for conducting a training impact analysis were developed and implemented within the context of an Operational Test (OT) of three medium antitank weapon systems and an Advanced Concept Technology Demonstration (ACTD) of 116 off-the-shelf technologies for urban operations. Data collected were predominately observational, consisting of time-referenced specimen records recorded sequentially within their naturally occurring context (i.e., observers collected data passively *in situ*, without control over test or training procedures). Between two and four independent observers were used at any one time, depending upon the particular circumstances of each test.

Observational training impact data were used to identify and compare the tasks soldiers had to learn and perform with different candidate systems. Subjective judgments were made about the relative complexity and difficulty of tasks across systems. The estimation of task difficulty in the OT was also supported by the use of two analytic models. Relative to a baseline technology or predecessor system, each candidate was ultimately judged to have either a positive, neutral, or negative training impact on the training base. Training impact rankings of systems were based on the relative number of tasks involved, the relative complexity and difficulty of each task, and the relative levels of training resources needed to achieve operational proficiency.

Findings:

Training impact methods that emphasize the direct observation of training and performance can be adapted to vastly different research situations, as illustrated in this report. In both the OT and ACTD, training impact differences were found among some or all of the candidates for most operational system requirements. For some requirements, however, no differential training impact was noted across candidates. Overall, there appeared to be

comparatively fewer training impact differences among candidates for a single operational system requirement than there were among candidates across requirements. Although of little value in making selection decisions, training impact comparisons across requirements may be beneficial in alerting and focusing the efforts of the training development community to those systems likely to have the most negative training impact when fielded.

Utilization of Findings:

In developing and proposing the use of training impact analysis methods, our experience has been that many of those making source selection decisions do not fully appreciate and understand the potential benefits of this kind of analysis, at least when the proposed methods are presented abstractly during an initial briefing. However, after they have seen the results of an actual analysis, many appear to be quite enthusiastic about its value for making more informed selection decisions. It is hoped that this report will not only guide other training developers and analysts in the conduct of their own training impact analyses, but that it will help to better illustrate the rationale behind training impact analysis methods for decision makers. The field research tips cited here will help other researchers who collect data using direct observation, both during the day and at night.

Results of training impact analyses can be used for purposes other than making source selection decisions. For example, information on training impact can give training developers a head start in the design of training programs, devices, and materials prior to the acquisition and fielding of new systems. This kind of information can also help to identify deficiencies in system design, which can sometimes be easily corrected if identified early in the product development cycle. Finally, training impact information can help develop more accurate budget forecasts of the training resources likely to be expended during new system fielding.

DIRECT OBSERVATION IN THE CONDUCT OF TRAINING IMPACT ANALYSES

CONTENTS

LIST OF TABLES

DIRECT OBSERVATION IN THE CONDUCT OF TRAINING IMPACT ANALYSES

Introduction

Procedures for conducting a training impact analysis were developed to provide timely information to those responsible for making selection decisions in the context of military testing and evaluation programs. Findings from a training impact analysis can help decision makers understand the training implications associated with selecting one particular candidate over others, although training impact is only one of many factors to be considered in the overall selection process. A training impact analysis forecasts or estimates the overall impact to institutional and unit training that a candidate system would have if it was selected for acquisition and fielding. Positive training impacts occur when a new system leads to an overall reduction in requirements for training resources. This can happen, for example, when a new system having a minimal training burden replaces an existing system with a heavier training burden. Unfortunately, neutral or negative training impacts appear to be more common than positive ones. Training impacts are neutral in situations where the overall training burden does not change following the introduction of a new system, even though the specific training tasks involved may change. Negative impacts occur when a new system increases the training burden overall, especially when a new system is merely added to an existing inventory of systems or when no predecessor system exists (i.e., all new tasks become pure additions to the training load).

Training impact analysis methods are good choices for training developers and training researchers to use in testing situations where they can only observe, but cannot control, the conduct of training and test events. Although less precise and conclusive than a formal training effectiveness analysis (Department of the Army, 1994), a training impact analysis allows selection decisions to be informed by training implications early in the product development cycle. Because training impact results are largely ordinal in nature, they are best used in comparative evaluations involving more than one candidate for a particular operational requirement. Candidate systems can be compared to other candidates or to a baseline or predecessor system currently in use. Without considering their operational effectiveness, candidates can be rank ordered in terms of their potential training impact to the extent meaningful variation exists in measures like the number of tasks to be trained, the complexity or difficulty associated with training and performing each task, recurring training costs, and special training requirements (e.g., the need for new or highly specialized training facilities).

Training impact methods can be tailored for the specific needs associated with a wide variety of research situations. As evidence of that assertion, the present report describes the conduct of training impact analyses within two highly distinct contexts. Initially, we describe an in-depth analysis conducted in conjunction with an Operational Test (OT) of three medium antitank weapon systems. Subsequently, we describe a series of relatively broad analyses conducted in conjunction with an Advanced Concept Technology Demonstration (ACTD) program. In this particular ACTD program, different kinds of commercial and governmental off-the-shelf systems were evaluated for their potential use in urban military operations.

Our procedural model for conducting a training impact analysis is composed of four general steps. First, obtain available information on candidate and predecessor systems. Detailed system information, if available, will allow a preliminary task analysis to be conducted and will enable the development of a training impact data collection plan. Second, observe training on each system *in situ*. The direct observation of system use in the hands of soldiers forms the core raw data upon which subsequent analyses are performed. In some situations, it may also be possible to observe the operational use of candidate systems in a force-on-force tactical training environment, where it is not unusual for soldiers to encounter problems not addressed in previous training. Third, examine the observational records obtained in the second step for evidence of task variation across candidate systems and estimate the relative difficulty or complexity of performing each differentiating task on each applicable system. Some tasks may not be performed with some candidates, based on their particular design features and operational characteristics. Collectively, we refer to all activities conducted in this third step as comparative task analysis, though a variety of distinct analytical methods can be used. Fourth, rank order systems in terms of estimated training impact, based on the results of analyses performed in the third step. Time permitting, one can also develop alternative programs of instruction (POIs) to verify rankings and to more clearly define the precise areas in which differential training impact exists across systems. Unlike many of the training estimation models cited by Muckler and Finley (1994a, 1994b), our approach did not measure training effectiveness, estimate training costs, identify training device requirements, estimate personnel requirements, or apply training media and instructional delivery models in any formal fashion.

The ability to gather system information varied between the OT and ACTD. In the OT there was a clearly defined predecessor system on which substantial training analysis, training resource, and test information were available. This allowed a preliminary task analysis to be performed and a training impact data collection plan to be developed. In the OT there was also more documentation of candidate systems, longer turn-around time for feedback to decision makers, and sufficient time to observe predecessor system training in both institutional and unit settings. In contrast, the ACTD involved a much greater number of candidate systems, for which comparable predecessor systems did not always exist. Typically, a brief "technology profile" of one or two pages was the only information that could be obtained prior to the first day of observational data collection. On occasion, a demonstration of a candidate system given by a manufacturer's representative could be witnessed prior to data collection.

Techniques of observational data collection, which were central to our efforts in both the OT and ACTD, have long been used in educational and social science research. Our observational records of training were semi-structured, involving neither the use of unstructured participant diaries nor highly structured checklists of categorized behaviors. Instead, the approach we used appears similar to the concept of collecting specimen records, in which one attempts to describe behavior sequentially within its original context (Bickman, 1976). Behavioral descriptions contained in specimen records tend to be continuous over a period of time, rather than being sampled, and they can include inferences made by observers. Evertson and Green (1986) would probably classify our general approach as a narrative system based on specimen records. However, it was not exclusively observational, as we felt free to ask occasional questions of both soldiers and instructors, primarily to confirm or refute inferences contained in our observational record of events. A major advantage of observational approaches

is that they generally tend to be less intrusive than formal interviews or questionnaires. Yet, our observations were not entirely unobtrusive. Research participants knew we were observing them and it was not uncommon for some to give us their unsolicited opinions of the systems being examined, as our constant note taking was conspicuous.

In both the OT and ACTD, military personnel had either little or no prior experience with the systems being compared. Thus, no pool of subject matter experts (SMEs) existed to be surveyed or interviewed in order to determine system requirements or estimate task complexity. In comparison, many training estimation models and techniques assume the availability of a large pool of subject matter experts for this purpose (see Muckler & Finley, 1994a, 1994b). As a result, our training observations were often the primary means of obtaining information on the full scope of system tasks, the ease or difficulty with which soldiers performed them during training, and the degree to which tasks performed during training mirrored tasks performed during the operational employment of the system in a tactical setting. Only in the ACTD, however, were we able to observe system use in the context of tactical exercises.

Critical to estimating the relative training impact of candidate systems is whether required tasks vary in difficulty or complexity from the viewpoint of the user, that is, the difficulty of performing system tasks (Meister, 1999). The techniques we used to derive these estimates have elements in common with training models cited by Muckler and Finley (1994a, 1994b), such as comparative task analyses and front-end analyses. However, we found formal observations of soldiers being trained on each system to be essential. On-site observations of training are typically not stressed as elements in other training estimation models (see Mucker & Finley, 1994a, 1994b), but were a crucial part of the approach presented here. Yet, our observational records were essential elements in analyzing system tasks and estimating task complexity or difficulty. In addition, where predecessor system training could be observed, on-site observations helped provide a realistic and accurate picture of baseline training resource requirements.

Understanding and estimating relative levels of task complexity are perhaps the most important and most difficult aspects of training impact methodology to perform. Many factors were used to estimate task complexity in both the OT and ACTD. These included the number of discrete steps in a task, whether or not steps must be performed in a particular sequence, whether or not some actions are contingent upon the occurrence of a particular set of conditions, the difficulty of recalling procedures in the absence of regularly occurring practice sessions, the level of performance feedback provided (i.e., knowledge of results), and the degree to which previously learned tasks enhance or interfere with the acquisition of new skills and knowledge. Those who attempt to conduct a training impact analysis must be able to observe the use of new systems *in situ* and be able to conceptually translate those observations into a corresponding sequence of human performance tasks. Comparisons of relative task complexity among candidate systems can then be made using factors such as those mentioned in this paragraph. Not only is it helpful for training impact practitioners to understand task analysis methods generally (see Drury, Paramore, Van Cott, Grey, & Corlett, 1987; McCormick, 1976; Meister, 1985), but it is imperative they be able to gauge the relative complexity of highly disparate tasks (e.g., the difficulty of Candidate A's Task 3C compared to Candidate B's Task 7D). More time

was available for task analysis activities in the OT than in the ACTD. For that reason, skill retention and cognitive requirements models were also applied to critical system tasks in the OT.

The techniques used to rank candidate systems were generally similar in both the OT and ACTD. However, sufficient time existed in the OT context to develop a series of alternative POIs that supported the rankings by more clearly defining the precise areas in which differential training impact existed across candidate systems. Unlike the OT, which involved one training impact analysis, the ACTD involved a series of training impact analyses. This permitted our general approach to be used with many kinds of candidate systems, differing widely in their design characteristics and operational complexity (e.g., from a kneepad for joint protection to an unmanned aerial vehicle for intelligence collection and dissemination).

Findings from the training impact analyses reported herein have been distributed only on a very limited basis, primarily to those charged with making candidate selection decisions on these two projects. For purposes of illustration, the present report offers excerpts of our training impact analysis methods and findings from both the OT and ACTD. However, we purposely do not refer to the actual names of any candidate systems. This is being done to focus on the research methods themselves and to insure this report can receive the widest possible distribution, by not referring to potentially sensitive evaluation information.

Because more training impact observations are made in the field than in the classroom, we present a series of field research tips in Appendix A that may be of interest to a wider audience, especially to those who are new to the area of applied military research. These "tricks of the trade" have been humbly learned in trial-and-error fashion by the authors over the course of their careers in applied military training and education research. Though practical and immensely logical, these relatively simple tips are really the kind of stuff that nobody teaches in textbooks or in graduate school.

Training Impact Analysis in an Operational Test

The training impact analysis described here was performed in the context of an operational test that compared three candidate antitank systems to a predecessor system (Dyer, Lucariello, & Heller, 1988). The predecessor system had been in the field for approximately fifteen years, thus providing a definitive baseline system. The research and development phase for the candidate systems provided the time and opportunity to obtain training data on these candidates as well as on the predecessor. In addition, there was time to conduct additional training analyses after the training for the operational test was completed. The analytic work covered a period of two years. For estimating the relative impact of candidate systems, we ordered them on difficulty of training, difficulty of performing, and resources required to train.

Approach

Overview of the analysis. We observed training on the predecessor system, first in the institution and then in the unit. Task analyses were conducted on all predecessor system tasks, specifying the steps in each task and the knowledge and skills required to complete each task.

Reports of predecessor system tests were reviewed. These analytic efforts preceded work with the candidate systems.

Next, documents describing the candidate systems and general system requirements, to include the training device requirements document, were reviewed. On the basis of these documents, we were able to perform initial task analyses on each candidate. Observations of training on the candidate systems occurred when soldiers were trained prior to the OT. These observations provided a check on the completeness and accuracy of our task analyses as well as information on task difficulty and complexity. The actual operational testing of the candidates was not observed.

All efforts culminated in estimating the relative training impact of the candidates to each other and to the predecessor. For estimating the relative impact of candidate systems, we ordered them on difficulty of training, difficulty of performing, and resources required for training. Skill retention and cognitive requirements models, applied to critical tasks on the predecessor system and each candidate, provided additional estimates of task difficulty. Lastly, we developed two alternative programs of instruction (POIs) for each candidate, which represented hypothetical upper and lower bounds on training resources and time.

Task analyses. Task analyses were first conducted on the predecessor system. Existing documentation, soldier's manuals, POIs, and test reports were the starting points for this analysis. The task information in this documentation was supplemented by field observations of training and interviews with instructors. The analyses included specifying knowledge and information requirements, and special skills. All task steps and decision-points were documented. The training observations and interviews with instructors were critical as they often showed that existing documentation of tasks was incomplete.

A similar process was followed for the task analyses of the candidate systems. However, in this case, the only documentation of task requirements was that from the equipment manufacturer.

Field observations. We designed a structured data collection form to record our observations of training, including platform instruction, practical exercises, and performance tests (Appendix B). The form provided a mechanism for generating a comprehensive record of the training. It also provided a means for recording each observer's immediate assessment of overall training quality, observed training problems, potential training improvements, task sequencing, and the knowledge, skills, and abilities required by soldiers to perform required tasks.

The raw data forming the basis of these observations were sequential narrative descriptions of events observed during the training. Events included individual behaviors, key teaching points of the instructors, soldier questions, practical hands-on exercises, an end-of-block test, preparation of training devices for firing, and instructor demonstrations. The narrative description or specimen record of each training event was time referenced. The time records, specifically start and stop times read from a digital watch, were summarized by training event or task. Whenever soldiers practiced tasks, performance times for each individual were recorded.

For many tasks, accurate times on the candidates were needed in order to assess the candidates against the time standards used with the predecessor.

We initially observed training on the predecessor system in institutional and unit settings. In these training settings there were three observers. Having multiple observers provided a check on the accuracy and completeness of the data. Each observer was assigned to a different group of students rotating through a series of training stations (i.e., round-robin training). However, each observer was assigned to a different instructor whenever one-on-one, instructor-to-student training occurred (e.g., with training devices used to simulate missile firing). Separate data records from each observer were then combined to obtain a total picture of the training.

We had four observers during the training of the three candidate systems. One observer was assigned to each candidate as the primary data collector for that system. The fourth observer "floated" from candidate to candidate, collecting formal data during each visit and providing quality control (i.e., a reliability check on the other observers). Because candidate observations occurred after the predecessor observations, observers were well-trained in our observation procedures when candidate training occurred.

We found it essential to write-up our observation notes, or at least review them for completeness and clarity, at the end of each day. If an event or action was not clearly described in our notes, it was then clarified the next day with the instructors or the soldiers. Observers must be attentive at all times to the actions in a classroom or field setting. Often we would use "short-hand" to indicate what happened. We learned very quickly that unless these abbreviated notes were translated and expanded shortly after the event, much of the critical detail was forgotten. One of the difficulties with this type of training observation is that one cannot prejudge what events are critical. Something thought to be incidental at the time often turns out to be critical later. Thus it is important to record as much as possible, without judging the material per se.

Test reports on predecessor. The analysis integrated existing test data on the predecessor system. Many extensive training and operational tests, to include tests of training devices, had been conducted with the predecessor. A thorough review of all these test reports was conducted. The reports contained "hard" performance data, e.g., results from missile firing, providing quantitative indices of task difficulty. The results provided an excellent background on the scope and difficulty of the predecessor tasks, as well as associated training resources. Not only could we compare the candidates to each other, we could also compare them to the predecessor. In particular, we knew which predecessor tasks were very difficult, and thus were mindful of these when observing the candidates. Consequently, a critical aspect of our analysis was to determine whether these difficult predecessor tasks were easier to train and to learn with any of the candidate systems.

Analytic models. Judgements made by the observers, plus the output from two analytic models, were used to assess the relative difficulty of the tasks on each system. On the basis of this analysis, a task fell into one of two categories. In one category, a task had about the same level of difficulty on each system. In the other category, a task differed in difficulty across systems. The observers' judgements were applied to all tasks and were based on training times,

the skill and knowledge analyses of each task, and soldier performance during training (e.g., attempts needed to perform the task correctly, errors made, or probabilities of target detection). In contrast, the analytic models were only applied to two critical tasks (i.e., prepare for firing and engage targets).

One of the analytic models we used was a skill retention model (Rose, Czarnoleswski, Gragg, Austin, Ford, Doyle, & Hagman, 1985; Rose, Radtke, Shettel, & Hagman, 1985). The other was a cognitive requirements model (Rossmeissl & Alderson, 1987). The skill retention model is based on ten rating scales that focus on such factors as job aids, number of task steps, built-in feedback from step to step, and number of facts. It yields a single score that corresponds to a particular 12-month decay curve, reflecting the predicted percentage of individuals able to perform the tasks correctly after given periods of time without intervening practice. The decay curves are valid only for tasks mastered during initial training. The cognitive requirements model also generates a single score that reflects the cognitive demand of the task. It is based on such dimensions as working memory, long term memory, quantity of data, problem solving, and multiple levels of processing. A high cognitive demand for a particular task is determined by comparing the score on that task with scores on other tasks on the same system, with scores on tasks performed by individuals in the same duty position, and/or with scores on the same or similar tasks on another system.

Both models were determined to be more appropriate for representing the demands made by tasks stressing procedural knowledge, i.e., knowing how (Anderson, 1980), than tasks stressing psychomotor and sensory integration (Fitts, 1964). Although the distinction between motor and cognitive processes does not always serve a useful purpose (Fitts, 1964), we viewed this distinction as important in characterizing the predecessor system tasks. As shown by extensive system tests and live-fire data, the psychomotor and sensory skills required to hit targets with the predecessor were very demanding. A brief description of these skills follows. The soldier must maintain a steady position and keep his eyes open while the missile is launched from his shoulder. This launch generates considerable heat that can burn the soldier, a noise level of 178 decibels, and debris which obscures the soldier's vision for the first 2-5 seconds of missile flight. Assuming the launch is successful, the soldier must continue to aim at the target, not at the infrared source on the missile, and hold his breath during the missile's flight down range. Body movement must be minimized as this movement is transferred to the missile. The soldier can easily ground the missile because of this movement or consume all the missile's thrusters in attempting to get the missile back on course, resulting in the missile falling short of the target. Test data showed that many soldiers failed to hit the target because they grounded the missile immediately after launch or tracked the missile rather than the target. The two analytic models did not provide a means of representing the difficulty associated with this task. The result was that they underestimated the difficulty of the target engagement task with the predecessor. With the candidates, however, engineering design efforts reduced the psychomotor and sensory demands required during target engagement. Consequently, the skill retention and cognitive requirements models more accurately depicted task difficulty with the candidates.

POI development. To obtain a more complete picture of the candidates' training impacts, we suggested two alternative POIs for each. We based these POIs on our training observations of the candidates, resources used to train the predecessor in the institution, principles of

instructional design, and the proponent's training goals. These POIs reflected a lower bound and an upper bound on training resources. The POIs were then compared to the predecessor POI as well as to an enhanced predecessor POI, which we designed to better reflect the full range of tasks and skills on that system.

Examples of Techniques and Findings

Task complexity. We compared systems by examining the demands placed on soldiers by the various tasks. In this report, we illustrate this approach with the malfunction procedures task. Malfunction procedures must be easy to execute and remember because they are done under stress and are performed infrequently. Hangfires and misfires are rare, but when they occur the soldier must make the right decisions and take the right actions very quickly. The fewer steps and fewer decisions, the better. For example, we interviewed a soldier who had a malfunction with the predecessor system, where a wrong decision or action on his part could have resulted in serious injury to himself and others. Even though the malfunction had occurred several years prior to the interview, the soldier clearly remembered the event and the stress he was under at the time.

Table 1 presents the factors considered in comparing systems on the complexity of their malfunction procedures. The more factors the soldier must consider, the more complex the procedure. The more steps or cues, the more complex the procedure. The steps involved in determining malfunction actions were represented in a decision-tree format. Some systems required soldiers to make a series of if-then decisions to determine what the next action should be. The number of decision points varied with the system. The cues that triggered those decisions varied across systems as well. Lastly, some systems required soldiers to continue a certain action for a specified time before proceeding to the next step. As shown in Table 1, Candidate A had the simplest malfunction procedures. It ranked as being the least complicated on half the six dimensions shown in Table 1. No other candidate had that advantage, with Candidate C being the least complex on only one dimension.

Training time and mode of instruction. In addition to comparing systems on individual tasks, our observations of the instruction given on the candidate and predecessor systems were condensed into data summaries that specified times allocated to each task and mode of instruction. Summaries of the training and administrative times for each system were generated for each day. Table 2 shows these summary data for one candidate. We also generated more detailed summaries of the time required for each block of instruction on each day. Table 3 reflects the same training as that in summarized in Table 2, but is organized in terms of the instructional methods and times devoted to each task.

The form in Appendix B was used by observers to collect the data presented in Tables 2 and 3. The form also prompted observers to provide extensive detail on the task requirements for each candidate: how each task had to be performed, the controls on each system, how the controls interacted with each other, decisions that had to be made by the soldier, what soldiers learned from the training devices, and what worked and what did not work during training. Although these findings are not summarized here, they constituted the core of our training impact analyses and constituted the rationale for the relative impact of each system (Dyer,

Table 1
Estimating the Complexity of Malfunction Procedures Across Systems

	Predecessor	Candidate A	Candidate B	Candidate C
Types of malfunctions	2	2	2	2
Decision points	5	0	1	5
Terminal branches	5	1	3	6
Steps in longest branch	4	6	5	3
Cues that trigger decisions	2	1	2	2
Time requirement for at least one action (e.g., squeeze trigger for x seconds)	Yes	Yes	No	No

Table 2
Example of a Training Time Summary for a Candidate System

	Time (minutes)		
Training Phase and Day	Instruction	Administration	Total
Initial Classroom Instruction			
Day 1	322	115	437
Day 2	227	116	343
Day 3	279	70	349
Day 4 (half day)	55	72	373
Range Training			
Day 5	37	91	128
Day 6	93	83	176
Total minutes	1013	547	1560
Percent of total time	65%	35%	100%

Table 3
Example of a Time and Instructional Methods Summary by Tasks and Training Events for a Candidate System

Tasks and Training Events	Time (minutes)	Time (% of total time)	Instructional Methods and Media
Tasks[a]			
Engage targets	703	69.4%	Lecture, gunner handbook,
System display & controls	(296)	(29.2%)	computer displays, computer tutorial & quiz. Video tapes of imagery
Thermal imagery	(204)	(20.1%)	PE with mockup & computer
Practice with mock-up	(154)	(15.2%)	tutorial; prototype system on range with vehicle targets. PE with system mock-up
Firing positions	(26)	(2.6%)	Lecture with viewgraphs
Surveillance	(14)	(1.4%)	Lecture with viewgraphs;
Determine if target is in range	(9)	(0.9%)	gunner handbook
Maintain system	57	5.6%	Lecture, gunner handbook
Carry system	45	4.4%	Lecture with viewgraphs; instructor demonstration
Prepare system for firing and Restore to carry configuration	22	2.2%	Instructor demonstration & gunner practical exercise (PE) with mockup
Perform malfunctions	8	0.8%	Lecture
Construct fighting position	0	0.0%	---
Prepare antiarmor range card	0	0.0%	---
Decontaminate	0	0.0%	---
Destroy system	0	0.0%	---
Training Events			
System information & background	125	12.5%	Lecture with viewgraphs
Test procedures	53	5.0%	Lecture with viewgraphs
Total time	1013	99.9%	

[a] Tasks ordered by amount of training time.

Lucariello, & Heller, 1989). In particular, our analyses sought to explain why certain tasks or steps on one system were more difficult than those on another system, why proposed training devices could or could not meet a training requirement, which tasks were the most critical and required a high degree of proficiency, and where practical exercises were needed. The descriptive information obtained in any training impact analysis that includes formal

observations of training may, in the long run, be of paramount importance, as it provides the foundation for system rankings and training recommendations.

Estimating Training Impact

Task difficulty. Of the ten primary tasks cited in Table 3, five were found to be highly similar across systems in terms of training impact (e.g., estimated difficulty, resource requirements, and amount of training needed). These tasks were: construct a fighting position, prepare an antiarmor range card, carry the system, decontaminate the system, and destroy the system in an emergency situation. In contrast, the other five tasks were found to differ across systems. These latter tasks were: maintain the system, prepare the system for firing, engage targets, conduct malfunction procedures, and restore the system to its carrying position. Table 4 summarizes the relative difficulty of the tasks judged to have a training impact. These five tasks included those that were considered the more critical system tasks, specifically, engage targets and conduct malfunctions. Engage targets is the primary system task as the purpose of the system is to engage or "kill" tanks. The malfunctions task is also critical because if malfunction procedures are performed incorrectly, soldiers can be seriously injured.

Table 4
Relative Difficulty of Tasks Across Systems From the Easiest to the Most Difficult

Engage Targets	Maintain	Prepare to Fire	Restore to Carry Position	Malfunction Procedures
Candidate A	Candidate A	Candidate A	Candidate A	Candidate A
Candidate B	Candidate B	Candidates B & C	Candidate B	Candidate B
Candidate C	Candidate C	Predecessor	Predecessor	Candidate C
Predecessor	Predecessor		Candidate C	Predecessor

The target engagement task received special attention, as it had been shown historically to be the most difficult of all predecessor tasks. A skill and knowledge analysis of the systems found overlap in the skills required to engage targets across systems, but this analysis also showed that each candidate system required unique skills. A layout of the firing sequence for each system also showed alternative firing procedures for each (i.e., there was more than one way of firing a missile). For comparison purposes, we defined a "quick fire" sequence and a "decision making" sequence for each system. The skill retention and cognitive requirements models were then applied to each sequence. Greater differences among systems were found with the "decision making" sequence as opposed to the "quick-fire" sequence. Nevertheless, the results of each model agreed with the task difficulty rankings in Table 4, irrespective of firing sequence.

POI development. Our training impact analysis of the antitank weapon systems went beyond the ordering of candidate and the predecessor systems solely on the dimension of task difficulty. We also designed institutional POIs for each system in order to determine and compare their respective total resource requirements. In developing these POIs, we suggested

training times, resources, device usage, task sequences, types of PEs, and instructional media. In making these suggestions, we followed general instructional design principles provided by Kyllonen and Alluisi (1987) on the learning and retention of facts and skills. Table 5 shows these key instructional design principles and our application of each to the development of the POIs.

As we developed POIs for the candidate systems, we felt it was necessary to also develop an enhanced POI for the predecessor. The goal of the system proponent was to train all skills and tasks on the candidates. However, training on the predecessor system did not encompass all requisite tasks and skills. For example, the night capability of the system, employment in a field environment, and maintenance of training devices were not trained in the courses we observed. Training media to support the predecessor's night firing capability and other aspects of target engagement did not exist. The enhanced POI that we developed covered all predecessor tasks, skills, and corresponding training resources. It provided a fairer basis for comparison with the candidates, though both the enhanced and current POIs were actually compared to the candidate POIs for the record.

We outlined two POIs for each candidate. The first POI closely paralleled the scope of the enhanced POI for the predecessor. It also adhered to the mission profiles established for the four training devices in the training device requirements document that was approved for all candidate systems. Differences among training times for the candidates were based primarily on the number of steps involved with different tasks, the ease with which soldiers had performed tasks during training, and a careful sequencing of basic to more complex target engagement skills with continual training of basic skills. This POI reflected an upper bound on training resources.

The second POI was a streamlined POI representing a lower bound on training resources. With this POI, only two of the four training devices were used. There was little redundant training of skills, making every block of instruction critical. Device training was reduced to only the essential skills. Several blocks of instruction were reduced to a lecture only, with no practical exercises. Table 6 summarizes the POI times for the alternatives examined in our analysis, including two different assumptions for the amount of administrative time required. Although the candidates ordered the same in terms of time to train regardless of the POI alternative, the absolute difference between the training hours for the candidates was greater with the more resource-intensive POI alternatives.

During our work with the antitank weapon systems it became clear that the training record itself is vital. What actually happened during training and why it happened were important issues. We soon became convinced our on-site, systematic documentation of soldier performance during training could provide critical information (e.g., typical errors made by soldiers or diagnostic skills required of trainers) to supplement formal training documents like POIs and training support packages. Such records can also provide greater insight about the performance of soldiers during system tests (Dyer, Lucariello, & Heller, 1989).

Table 5
Instructional Design Principles Applied to the POIs

Design Principle	Examples of Application of Design Principles to Programs of Instruction
Task-analyze the learning domain.	Task analysis conducted on each system.
Organize instructional goals around behavioral objectives: Instruction should be goal-oriented.	Instruction based on the Army's philosophy to "train as we will fight."
Show positive and negative instances of concepts.	Stressed in media and instruction on antiarmor range card, target acquisition, and target engagement.
Shape successive approximations to target performance. Provide extensive advice during early stages of learning, with less advice as learner becomes more skilled.	Applied primarily to time allowed for instructor feedback on the training devices developed for gunner target acquisition and weapon firing skills.
Minimize working-memory load.	Tasks sequenced to avoid information overload. For example, training on system controls progressed from basic engagement skills to skills required in less frequent situations.
Provide immediate feedback on error.	Feedback was a training device requirement.
Maintain active learner participation.	High level of practice on cognitive and motor skills stressed. Practice and test environments/ scenarios designed to simulate combat.
Maximize critical skills while training.	Resources and time allocated for critical gunner skills when firing weapon.
Achieve operational fidelity to promote generalization of skills.	Target engagement practice under a variety of conditions was stressed.
Train under mild speed stress.	Part of the practice trials on the devices required soldiers to perform tasks quickly.

Table 6
Time Comparisons of POIs

POI	Time in Hours with 20% Administrative Time	Time in Hours with 30% Administrative Time
Predecessor		
Current	30	34
Enhanced	68	78
Upper Bound for Candidates		
Candidate A	63	72
Candidate B	65	74
Candidate C	70	80
Lower Bound for Candidates		
Candidate A	35	40
Candidate B	37	43
Candidate C	39	44

Training Impact Analyses in an Advanced Concept Technology Demonstration

The Military Operations in Urban Terrain (MOUT) ACTD is a joint program of applied experimentation in the U.S. Army and U.S. Marine Corps that seeks to identify and evaluate the ability of various advanced off-the-shelf systems to meet specific needs associated with urban military operations. The identification and evaluation of candidate systems were guided by a set of 32 jointly developed and prioritized user requirements (e.g., the need for non-lethal stun grenades or the need to quickly get on top of buildings). A series of 10 separate MOUT ACTD experiments, six in the Army (AE1-AE6) and four in the Marine Corps (ME1-ME4), were conducted. Each of these experiments involved the evaluation of one or more candidate systems within the context of one or more of the 32 user requirements.

Each MOUT ACTD experiment within the Army consisted of five phases: initial training and certification, new equipment training (NET), side experimentation, tactical experimentation, and collective experimentation. The purpose of the first phase was to insure soldiers were able to perform important individual and collective skills in a simulated urban environment. Candidate systems were then introduced and trained. After NET, a series of precisely focused side experiments were performed, each attempting to answer a relatively narrow research question (e.g., how long it takes to assemble a candidate ladder). Tactical experimentation, consisting of a series of trials in which an Experimental Force (EXFOR) engaged an Opposing Force (OPFOR), was then conducted. Generally, each tactical trial focused on the effects of a single candidate system on EXFOR performance. After the best candidate for each user requirement was selected on the basis of side and tactical experiments, the selected systems were then examined as a package during collective experimentation.

Experimentation within the Marine Corps was organized into three phases, corresponding to the second, third, and fourth phases of Army experimentation. Although Marine Corps experiments tended to involve the participation of a greater number of individuals than did the Army experiments, all MOUT ACTD participants were drawn from infantry units within each service. Marine Corps tactical trials generally involved a series of EXFOR squads engaging a series of OPFOR teams. In contrast, each Army tactical and collective trial typically involved the same EXFOR platoon repeatedly engaging the same OPFOR squad. As a result of these organizational differences, Marine Corps trials tended to be more numerous and shorter in duration than Army trials.

Separate training impact analyses were conducted in conjunction with each of the 10 MOUT ACTD experiments. For each experiment, the differential training impacts of all candidate systems were compared. Candidates were either compared to all other candidates for a particular requirement or to a baseline system that already was being used in one or both of the two services. In general, a MOUT ACTD experiment took about three weeks to complete in the field. A report of the training impact analysis then had to be written and submitted within two to three weeks following the experiment's conclusion. All 10 experiments and their accompanying training impact analyses were conducted within an 18-month period.

The relatively high tempo of this experimentation had a direct bearing on the analytical approach chosen. It was not unusual for an experiment's schedule to change on a daily, or even hourly, basis. Thus, our approach needed to be as flexible as possible. After completing the training impact analysis report for one experiment, there was almost no planning time available before we had to launch the next training impact analysis. Although we had a vague notion of the kinds of systems we were likely to encounter in an upcoming experiment, we usually did not have an opportunity to see the candidates before the research participants themselves first saw them. Although this made planning more difficult, if not impossible, it did allow us to begin each experiment sharing a student's perspective with most research participants. If we had trouble understanding some of the instructional material upon hearing it for the first time, we assumed they did also.

Approach

The raw data forming the basis of each training impact analysis were sequential narrative descriptions of events observed during each phase of an experiment. Each narrative description, or specimen record, was time referenced using each observer's wristwatch to the nearest half minute. Although we initially used a digital stopwatch to record the duration of events to the nearest second, that approach was quickly abandoned. Recording event duration precisely was found to be impractical for six reasons. First, its greater level of precision was not needed (i.e., most events lasted several minutes or more). Second, glancing at a wristwatch was more easily accomplished, given that one hand already held a notepad and the other held a pen. Third, the simple entry of start and stop times for discrete events, or start times only for each step in a continuous chain of events, was found to be much quicker than trying to calculate the duration of events as they happened (i.e., from stop and start times we could calculate the duration of events afterwards). Fourth, the wristwatch method automatically gave us records keyed to the time of

day (or night) that events occurred. Fifth, we found using a stopwatch distracted us from those we were observing, posing a distinct possibility of losing important data. Finally, using a stopwatch may unintentionally suggest to research participants that speed is an important factor, which is something observers should avoid.

Events could be either individual behaviors, group actions, key teaching points of instructors, student questions, a portion of a classroom period of instruction, an informal outdoor rehearsal session, an experimental trial, a radio transmission, or an after-action review (AAR), to name a few. From one to three observers recorded each event. In most cases, classroom training was presented to research participants as a single group. All available observers (usually two) were used in those situations. In contrast, research participants were typically divided into two or more groups during hands-on training sessions. For those situations, each available observer (usually two) was assigned to one of the groups. At times, not all groups could be observed. If groups rotated through a series of concurrent training stations, each observer remained with the original group to which they were assigned. On occasions when three observers were present, the third observer "floated" between concurrent stations to obtain a different perspective (i.e., by comparing the actions of two or more groups at each station). Although the qualitative nature of the narrative data precluded the measurement of inter-rater reliability, the observers met on a daily basis to compare notes and plan for the next day's events. It was not unusual for observational errors to be corrected in this manner, particularly when one observer in the group either recorded an improbable time entry or made an inference unsubstantiated by available fact. In instances where disputes could not be immediately resolved, we sought the input of others to eliminate the confusion (primarily instructors and those conducting the side experiments, as well as some research participants and other on-site witnesses to a lesser extent).

An excerpt from the raw data obtained by one observer during the first Army experiment (AE1) is shown in Table 7. The level of detail contained in these narrative descriptions is representative of what we obtained across the 10 experiments. Basically, all observers tried to record as much information as they could within the constraints of the experimental conditions. One must understand the experimental events did not stop just so observers could make precise notations. It was generally unwise to keep one's head buried in a notebook, as some important event could be missed (e.g., an EXFOR team is preparing to breach a building next to your location, putting you in immediate danger of being shot with a training munition if you don't move right away). Thus, the competing demands of direct observation and actual note taking had to be constantly balanced.

Table 7 also indirectly demonstrates that we did not really focus on how well the research participants performed. Although we had a general sense of overall performance levels, personnel from other agencies were directly tasked with measuring performance and effectiveness variables. However, we did note the general level of training performance with each candidate system during NET, recording most participant questions and the types of operational errors encountered. For example, we noted the frequent occurrence of accidental discharges with the Candidate B system for tactical engagement simulation, subsequent to the events depicted in Table 7.

Table 7
A Sample of Raw Training Impact Analysis Data

0855 hrs.	Arrive at Lane B of McKenna Range
	All soldiers in EXFOR squad present (n = 11)
	Soldiers loading training ammo (Candidate A for tactical engagement simulation)
	Wearing Candidate C knee & elbow pads on shins and forearms
	Wearing Candidate A knee & elbow pads on knees and elbows
0904	Issue Candidate B equipment for tactical engagement simulation
0919	Begin zeroing this equipment
0944	End zeroing
	PFC tells me "attacking the same buildings over & over again gets monotonous"
	Set-up threat silhouette targets with Candidate B sensors inside buildings
	Team A (n = 4) &Team B (n = 5) go to stack positions behind Bldg. 1; machinegun support team (n = 2) sets up in woodline
	All soldiers using both Candidates A and B for tactical engagement simulation
0957	Breach Bldg. 1 with artillery simulator; both teams enter through mousehole
0959	Team A moves across street to Bldg. 2 and breaches with artillery simulator
1000	Team B moves to Bldg. 2 and enters
1002	Team A moves to Bldg. 3 and stacks near rear door
1002.5	Team A breaches Bldg. 3 with an artillery simulator and enters through doorway
1003	Team B enters Bldg. 3
1004	2 soldiers clear stairwell with artillery simulator
1005.5	Cease fire; EXFOR regroups for informal AAR with SL
	Instructor critiques movement across open areas (i.e., crossing street one at a time)
1020	Informal AAR complete

On occasion, the time-referenced recording of experimental events allowed us to calculate the actual level of training throughput achieved over the course of an entire experiment. Table 8 summarizes the AE1 training throughput achieved with various combinations of baseline and candidate systems for tactical engagement simulation in a MOUT training environment. Readers should not bother judging whether one candidate system was better than another, as we have not presented enough information to do so here. Rather, the information in Table 8 is only offered as one example of an analysis that can be performed using raw training impact data.

Table 8
Throughput Associated With Six Different Training Conditions Involving Tactical Engagement Simulation Systems

Training Condition	Buildings Cleared per Training Hour
1. Baseline	4.20
2. Candidate A	2.50
3. Candidate B & Baseline	1.07
4. Candidates A & B	1.76
5. Candidates A, C, & Baseline	0.73
6. Candidates A, B, & C	2.67

Note. Over the six AE1 training conditions, 150 buildings were cleared at a rate of 1.75 buildings per training hour.

Nothing presented in this report can be truly prescriptive for every situation, as there were real differences among the analytic methods used in each of the 10 MOUT ACTD experiments. For that reason, illustrative examples are presented in the next section to highlight some of those differences. Still, each analysis was based on raw observational data like those shown in Table 7, with comparisons among the baseline and candidate systems generally made on the basis of the task requirement differences associated with training individuals and groups to use each candidate effectively. After presenting a series of illustrative examples of training impact techniques and findings obtained from the 10 experiments, we will discuss the way in which we arrived at overall estimates or ratings of training impact.

Examples of Techniques and Findings

Number of training tasks. Tasks that most influenced the relative training impact of candidate powered optic systems in the third Marine Corps MOUT ACTD experiment (ME3) are shown in Table 9. Training impact, overall, appeared to result more from the design characteristics of the candidates themselves than from their interface with particular weapon systems. Tasks shared by all systems (e.g., mounting and zeroing) were not included in Table 9.

Both in terms of their design features and their subsequent training impact, there were three broad classes of powered optic systems in ME3: unmagnified aiming points, magnified optics with fairly complex reticles, and thermal sighting systems. The estimated training impact of the unmagnified aiming point systems (Candidates A and B) was negative, but small. Aiming with these systems was practically self-evident. In fact, if such systems were made an integral part of a future rifle, eliminating iron sights, their training impact could be almost neutral (i.e., one could eliminate teaching sight alignment, but would have to teach either battery installation or tritium safety).

Table 9
Tasks Associated With Eight Candidate Optical Systems

Tasks	Candidate Optical Systems							
	A	B	C	D	E	F	G	H
Install battery and turn on/off	X	•	•	•	X	•	X	X
Adjust brightness	X	•	•	•	X	•	X	X
Adjust contrast	•	•	•	•	•	•	X	X
Adjust focus	•	•	•	•	•	•	X	X
Select polarity	•	•	•	•	•	•	X	X
Select field of view	•	•	•	•	•	•	X	•
Select reticle	•	•	•	•	•	•	X	X
Estimate range with reticle	•	•	X	X	X	X	X	X
Acquire magnified targets	•	•	X	X	X	X	X	X
Inspect tritium lamp	•	X	X	X	•	X	•	•
Employ with NVGs	X	X	X	X	X	X	•	•
Identify thermal images	•	•	•	•	•	•	X	X
Total number of tasks	3	2	4	4	5	4	10	9

Note. X indicates the presence and • indicates the absence of a task with a candidate system.

The magnified optics (Candidates C, D, E, and F) were estimated to have small to moderate negative training impact. Coincidentally, the reticles in these systems were more complex than those in the aiming point systems. In general, rapid target acquisition tends to be more difficult with magnified optics than with iron sights, especially in the absence of repeated practice. Small sight movements around a target that are barely perceived with iron sights (e.g., as a result of normal movement experienced in unsupported firing positions) can become quite disconcerting to the novice when they are magnified.

The thermal sights (Candidates G and H) appeared to have the greatest negative training impact (i.e., moderate to large) of any class of powered optics. In short, the operation of thermal sights is relatively complex and is currently foreign to most Marines who fight with small arms weapon systems. The images produced by thermal sights are also more difficult to interpret and identify than those produced in other systems. This is because the identification and interpretation of thermal images can have greater cognitive requirements. As suggested in Table 9, thermal sighting systems usually have a greater number of controls than other kinds of optical systems. Adjustment of these controls for optimum performance can also be more difficult to learn (e.g., adjusting brightness and contrast after selecting polarity and field of view).

Based solely on the number of tasks involved (see Table 9), estimated training impact differences within each class of powered optic systems were found to be much smaller than they were between classes. Considering differences both within and between classes, however,

Candidate B appeared to have the least negative training impact, followed closely by Candidate A. Greater negative training impact was associated with Candidates C, D, and F. These three candidates were virtually identical in terms of training impact, followed closely by Candidate E. Finally, the two thermal candidates appeared to have the greatest negative training impact, though the impact of Candidate H appeared to be slightly less than Candidate G.

Task complexity and training costs. The seven door and window breaching systems examined in ME1 were much more difficult to evaluate, perhaps because the breaching candidates had characteristics that were even more internally heterogeneous than the powered optic systems. For example, some breaching systems were explosive and some were non-explosive. Some required teamwork and some did not. Some had recurring training costs and some did not. Differences such as these led to heated discussion among the observers as to how best to rank the seven candidates in terms of training impact. To resolve observer disagreement, we decided to compare not only the tasks associated with each candidate, but also the relative level of complexity of each task. Using many of the task complexity factors previously mentioned, we found considerable variation among the candidates in terms of their total complexity (see Table 10).

Training costs were another factor influencing the impact of particular systems. Some types of costs could be applied across the board. For example, in order to properly master breaching techniques with these candidate systems, operators must have some amount of hands-on practice. Ideally, this would include the destruction of doors of several types (e.g., inward/outward opening, wood/metal construction, single/double width). The construction time and costs associated with door replacement are factors that must be considered for all of the candidates. Other types of costs would apply only to particular systems. For example, both Candidates E and F would have recurring training ammunition costs that the non-explosive devices would not have. This cost factor added to the relatively greater levels of negative training impact for these two candidates.

Considering both task complexity (see Table 10) and training costs, the seven candidate breaching systems were ranked in terms of their training impact as follows (from the least negative impact to the greatest negative impact):

1. Candidate A
2. Candidate B
3. Baseline and Candidate C
4. Candidates D and E
5. Candidate F

If the baseline system were replaced with either Candidates A or B, the overall training impact was judged to be positive. If the baseline were replaced with Candidate C, the impact was judged to be neutral. Replacing the baseline with any other candidate was judged to have a negative training impact.

Table 10
Estimated Complexity of Tasks Associated With Seven Breaching Systems

	Breaching Systems						
Tasks	BL	A	B	C	D	E	F
Employ as a team	2	•	1	•	1	•	•
Recognize different types of construction	1	1	1	1	1	•	1
Select best primary and secondary attack points for different types of construction	3	1	1	3	3	•	3
Perform maintenance	•	•	•	1	1	•	2
Perform immediate and remedial action	•	•	•	•	•	1	2
Operate hydraulic valve	•	•	•	1	1	•	•
Assemble and disassemble	•	•	•	•	1	1	2
Estimate range and adjust aiming point	•	•	•	•	•	3	•
Select ammunition	•	•	•	•	•	1	1
Engage targets with live fire	•	•	•	•	•	1	1
Total training complexity score	6	2	3	6	8	7	12

Note. BL = Baseline. The complexity of a task was estimated to be either low (1), moderate (2), high (3), or not applicable (•). Complexity scores were summed across applicable tasks to yield a total training complexity score. Higher total complexity scores indicated a greater level of negative training impact for these candidate systems.

Task complexity. We also estimated the complexity of five tasks associated with three ballistic shield systems examined in ME1, even though all observers had completely agreed on their ranking of these candidates beforehand. Estimates of training complexity are shown in Table 11. We concluded that Candidates A and B would have a small negative impact on training if adopted, but that Candidate C would likely have a moderate negative impact on training. The relative difficulty of offensive and defensive employment was mostly a function of shield size and weight, with larger and heavier shields being more difficult to use. Readers may note the level of detail shown is greater in Table 10 than in Table 11. That is true; the level of detail among tasks varied somewhat from analysis to analysis. However, in no experiment was the level of detail greater than that shown in Table 10. If no differential impact among candidates was found using that level of detail, we generally concluded there was no difference among the candidates for a particular user requirement.

Table 11
Estimated Complexity of Tasks Associated With Three Ballistic Shields

	Ballistic Shields		
Tasks	A	B	C
Employ offensively	1	2	3
Employ defensively	1	1	2
Operate light	1	1	1
Connect shields	•	•	2
Operate wheels	•	•	2
Total training complexity score	3	4	10

Note. The complexity of a task was estimated to be either low (1), moderate (2), high (3), or not applicable (•). Complexity scores were summed across applicable tasks to yield a total training complexity score. Higher total complexity scores indicated a greater level of negative training impact.

Training time and design characteristics. We found measurable training impact differences among the peripheral sets associated with five hands-free radio candidates in AE3, AE5, and AE6. Peripheral sets are plug-in devices that permit the wearer to send and receive radio messages while keeping both hands free for other important tasks (e.g., firing a rifle). Consider the differential characteristics of the peripheral sets shown in Table 12. In brief, we found a direct relationship between actual NET duration and the numbers of components and connections in peripheral sets. We concluded that negative training impact increased across hands-free radios in the same order as their associated peripheral sets are presented in Table 12, from the least negative to the most negative.

Nevertheless, any of the hands-free radio candidates would likely have a large negative training impact if adopted, with or without consideration of a peripheral set. Primarily, a negative training impact would be expected because most platoon members are not currently trained to use radios within the context of either squad or platoon operations. Although differences in training impact among the five candidates do exist, these differential impacts are likely to be dwarfed by the negative training impact caused by merely giving a radio to every platoon member. This negative impact has two origins. First, each soldier and Marine would have to learn the mechanics of radio and peripheral assembly, as well as the basic operational procedures associated with message transmission. Presumably, for greatest efficiency this training would be accomplished in basic training. However, current basic training procedures assume Privates only require an introduction to radio communication procedures prior to unit assignment. Indeed, some elementary principles are already taught during basic training, but the rehearsal time needed for complete skill development is limited. Second, current communication-related tactics, techniques, and procedures may be inadequate to manage the increased volume of message traffic caused by having radios throughout the platoon. While the methods needed to train individuals how to use radios are known, it is unclear how to best train a large group of soldiers to effectively communicate with each other when they are on the same network.

Table 12
Differences Among Five Hands-Free Radios and Associated Peripheral Sets

Radio	Peripheral Set Components	Separate Components	Required Connections	NET Time (minutes)
A	Headset with boom and bone microphones Ring push-to-talk switch	2	1	32
B	Headset with boom and bone microphones Palm push-to-talk switch	2	1	39
C	Handset Hanging earpiece	2	2	43
D	Interface box Ring push-to-talk switch Plug-type earpiece	3	3	54
E	Throat microphone Ring push-to-talk switch Hanging earpiece Interface box	4	6	61

Training issues other than impact. Sometimes we did not find training impact differences among the candidates for a particular user requirement. Such was the case in AE5 and AE6 with two explosive systems for breaching reinforced concrete walls. Although no differential training impact was found, we still had three important conclusions to make about training with these systems. First, an inert version of any selected candidate is needed for training. It is important that this inert trainer accurately replicate the size, weight, and flexibility of its live explosive counterpart. Second, a live training charge (i.e., with reduced explosive effects) is needed to simulate wall breaching in the context of squad mission rehearsals. Training charges also need to replicate the size, weight, and flexibility of their full explosive counterparts as much as possible. Finally, realistic target effects are critical for effective training. Essentially, live training charges need to create simulated man-sized holes in gypsum board that are similar in size and shape to the effects of full charges in reinforced concrete walls. One should also attempt to recreate the pattern of rubble that would result from an actual charge, using loose rock or pieces of concrete block in training exercises.

It was difficult to gauge the precise training impact of each explosive breaching system during AE5 and AE6. This was primarily because of the real potential for substantial recurring training costs, which cannot be accurately determined until a means of charge adhesion is selected and final versions of both inert and training charges for each candidate have been designed and manufactured. Nevertheless, it appeared either candidate would likely have at least a moderately negative training impact if adopted, even if not every soldier is taught how to

prepare, prime, and detonate these explosive devices. This is because the operational use of explosive devices must still be rehearsed within the context of squad wall-breaching missions.

Estimating Training Impact

The estimated training impact differences among candidate systems for a particular user requirement, illustrated in the previous section, were generally found to be much smaller than the relative differences among candidates across all requirements. In the former case, training impact estimates tended to take the following form:

> Candidate(s) <u>A and B</u> had <u>substantially/slightly</u> <u>less/more</u> <u>negative/positive</u> training impact than did Candidate(s) <u>C and D</u> for Reason(s) <u>1, 2, and 3</u>.

To the extent that at least some slight differences could be found, an ordinal ranking of the candidates for a particular requirement was made.

In the case of training impact comparisons made across requirements, the estimates were much easier to make and they tended to have a more definitive flavor. Compare the modifying terms used to describe training impact in the following three examples:

> All candidates for the X Requirement would have a <u>moderate to large negative</u> training impact if adopted.

> Any of the Y Requirement candidates are likely to have <u>some positive</u> training <u>impact</u> if selected for acquisition and fielding.

> Most of the candidates for the Z Requirement are likely to have no more than a <u>slightly negative</u> impact on training, in the worst possible case.

The primary reason the differences across requirements were larger and easier to evaluate was that the respective task sets among candidates across requirements categories were more heterogeneous than they were within requirements categories. Over the 18 months of MOUT ACTD experimentation, a seven-point scale of training impact gradually emerged that seemed to characterize the relative differences among systems across requirements. This impact scale is shown in Table 13.

Of the 113 candidate systems summarized in Table 13, there is a fairly wide distribution of systems across training impact categories. Yet, we found that most, if not all, of the candidates pertaining to a particular user requirement tended to fall in the same training impact category. For example, all candidate ladders (n = 5) appeared to have some positive training impact, while the impact of all candidate gloves for cut protection appeared neutral (n = 12). The average MOUT ACTD system, if there is truly such a thing, could probably be described as having at least a small negative impact on the training base.

Table 13
Frequency of MOUT ACTD Candidates in Each of Seven Training Impact Categories

Training Impact Category	Number of Candidates
1. Some amount of positive training impact	8
2. Neutral training impact (neither positive nor negative)	35
3. Small negative training impact	30
4. Small to moderate negative training impact	17
5. Moderate negative training impact	10
6. Moderate to large negative training impact	7
7. Large negative training impact	6

Note. Due to the need for additional investigation, three additional systems were not categorized.

Discussion and Conclusions

It is important to mention that a training impact analysis will probably, of necessity, involve subjective judgements on the part of analysts. The fact that subjectivity is required should not prevent one from making such judgements. In fact, those observers whose records form the basis of a training impact analysis may be the only individuals who have seen all the training. This alone probably puts them in the best position to make such determinations. Still, the critical part of this process is to insure that the rationale for these determinations is clearly described, so others can then make their own decisions about the validity of the judgements.

Compared with questionnaire and survey methods, training impact analyses can be manpower intensive. For example, observers must be on-site at all times. In addition, multiple observers are usually required in order to observe all events and to obtain all relevant performance data on a majority of the participating soldiers. Multiple observers also reduce the likelihood that bias will influence the findings and conclusions.

One of the challenging aspects of a training impact analysis is estimating task difficulty. The analyst must feel comfortable with the ratings and rankings assigned. As illustrated in this report, a variety of techniques can be used, depending on the circumstances. Obviously, the best source for such estimates is soldier performance in a test-like situation. But typically, those conducting a training impact analysis have little or no control over the instruction and testing procedures. Therefore, techniques for estimating task difficulty and task complexity can be valuable additions to the analytic arsenal. When possible, multiple procedures should be used to estimate task parameters. The outcomes of these multiple procedures can be compared before making final training impact determinations.

The illustrative examples of data, methods, and findings in this report suggest different ways to estimate the training impact of a candidate technology or system, depending upon the particular constraints of the research situation. Nevertheless, these examples all have one important element in common. Specifically, they are all derived from the direct, systematic, and time-referenced observation of soldiers and instructors in the field and classroom. The observational nature of training impact data collected *in situ* is what most differentiates it from questionnaire and survey data typically obtained from participants only at the conclusion of a training event.

There are four types of constraints encountered by researchers that most influence the selection of particular training impact analysis methods. The first is the amount of time available for observation. This is a joint function of the test design plan and the nature of the technology being tested, neither of which can typically be influenced by training impact analysts. Instead, one must adapt data collection plans to the overall test schedule, attempting to collect as much observational data as possible within the allotted time period. Second, safety and sensitivity issues surrounding some systems can sometimes restrict or limit opportunities for observation. Third, some technologies used by soldiers are just inherently difficult to observe. For example, soldier interaction with personal computer software for battalion staff planning was difficult to observe in the MOUT ACTD. While the interaction among staff members could be observed and the visual information on computer monitors could be observed if its size was not too small, we found it impossible to observe and record the series of computer keystrokes made by individual participants. In this situation, an analysis of operator errors could not be performed. The fourth major type of constraint to which researchers must adapt is the amount of time available for data analysis. This analysis time generally commences with the conclusion of data collection and ends with the preparation of a final report or briefing of findings. In this respect, the OT of antitank weapon systems afforded a more complete opportunity for analysis and interpretation than did the MOUT ACTD.

Training impact is but one factor, and rarely is it thought to be the most important factor, to consider in making final selection decisions among candidate systems for a particular user need or operational requirement. Certainly, selection decisions must also consider a host of other important factors such as military utility, system effectiveness, mobility, soldier acceptance, safety, and cost. However, the training impact of a candidate system can sometimes give it an edge over other candidates, especially in situations where most of the candidates are thought to be similar in terms of other selection factors. Although the candidate having the most positive or least negative training impact is not always selected for acquisition and fielding, our experience has been that decision makers consistently value and are highly appreciative of having training impact information to use in their deliberations.

The results of training impact analyses can also be used for other important purposes. For example, training impact information can give training developers a head start in the design of training programs, devices, and materials prior to the acquisition and fielding of new systems. Additionally, this kind of information can sometimes influence subsequent improvements to system design, because sources of negative training impact can often be traced to very specific design features or operational characteristics of a system. Finally, training impact information can help develop more accurate budget forecasts of the training resources likely to be expended

when new systems are fielded. If skill retention modeling is performed as part of an overall training impact analysis, one could also estimate the human performance decrements that would likely occur as a result of hypothetical training resource reductions of various magnitudes.

In conclusion, the use of direct observation in the conduct of training impact analyses appears to be highly adaptable to disparate research and test situations. Although the general guidelines presented in this report are meant to be illustrative, and certainly not prescriptive, we believe they can generalize to other types of military test and evaluation programs beyond OTs and ACTDs. While an observational approach to training impact analysis can be resource intensive compared to survey and interview methods, it appears to provide valuable training impact information to both decision makers and training developers early in the product development cycle.

References

Anderson, J. R. (1980). *Cognitive psychology and its implications.* San Francisco: W. H. Freeman.

Bickman, L. (1976). Observational methods. In C. Selltiz, L. S. Wrightsman, & S. W. Cook, *Research methods in social relations* (3rd ed., pp. 251-290). New York: Holt, Rinehart and Winston.

Department of the Army (1994, September). *The TRADOC training effectiveness analysis (TEA) system* (TRADOC Regulation 350-32). Fort Monroe, VA: Headquarters, U.S. Army Training and Doctrine Command.

Drury, C. G., Paramore, B., Van Cott, H. P., Grey, S. M., & Corlett, E. N. (1987). Task analysis. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 370-401). New York: Wiley.

Dyer, J. L., Lucariello, G., & Heller, F. (1989). AAWS-M training impact analysis: A case study. *MANPRINT Bulletin, 3*(6), 1-3.

Dyer, J. L., Lucariello, G., & Heller, F. H. (1988). *Advanced antitank weapon system-medium (AAWS-M) training impact analysis.* Fort Benning, GA: U.S. Army Research Institute for the Behavioral and Social Sciences Field Unit. [Competition Sensitive]

Evertson, C. M., & Green, J. L. (1986). Observation as inquiry and method. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 162-213). New York: Macmillan.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243-285). New York: Academic Press.

Kyllonen, P. C., & Alluisi, E. A. (1987). Learning and forgetting facts and skills. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 124-153). New York: Wiley.

McCormick, E. J. (1976). Job and task analysis. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 651-696). Chicago: Rand McNally.

Meister, D. (1985). *Behavioral analysis and measurement methods.* New York: Wiley.

Meister, D. (1999). *The history of human factors and ergonomics.* Mahwah, NJ: Lawrence Erlbaum.

Muckler, F. D., & Finley, D. L. (1994a). *Applying training system estimation models to Army training, Volume I; Analysis of the literature.* (ARL-TR 463). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Muckler, F. D., & Finley, D. L. (1994b). *Applying training system estimation models to Army training, Volume II; An annotated bibliography 1970-1990.* (ARL-TR 463). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., Ford, P., Doyle, J., & Hagman, J. D. (1985, February). *Acquistion and retention of soldiering skills.* (Technical Report 671). Alexandria, VA: U.S. Army Research for the Behavioral and Social Sciences. (AD A160 336)

Rose, A. M., Radtke, P. H., Shettel, H. H., & Hagman, J. D. (1985, July). *User's manual for predicting military task retention* (Research Product 85-26). Alexandria, VA: U.S. Army Research for the Behavioral and Social Sciences. (AD A163 710)

Rossmeissl, P. G., & Alderson, M. A. (1987). MANPRINT: A new design approach. *MANPRINT Bulletin*, *9*, 1-3.

List of Acronyms

AAR	After Action Review
ACTD	Advanced Concept Technology Demonstration
AE	Army Experiment (within MOUT ACTD)
BL	Baseline
EXFOR	Experimental Force
ME	Marine Corps Experiment (within MOUT ACTD)
MOUT	Military Operations in Urban Terrain
NET	New Equipment Training
OPFOR	Opposing Force
OT	Operational Test
PE	Practical Exercise
POI	Program of Instruction
SL	Squad Leader
SME	Subject Matter Expert

Appendix A

Field Research Tips

The following list of field research tips reflects both practices used during the conduct of our training impact analyses, as well as ongoing practices originated during previous field research projects. Though by no means exhaustive, the list was drawn from the personal experiences of the authors. Much of the material may appear rather simple or obvious, but hopefully it will be of benefit to new field researchers. However, we also hope that one or two tips may have some value to those more highly experienced.

Taking Notes at Night

One technique that can be used to minimize the negative effects of extraneous light on research participants at night is to do the following. With a knife or scissors, make a clean cut along one end of a foil chemlight wrapper. Remove and activate the chemlight. Then put the chemlight back into the wrapper, thick end first. Finally, insert a ballpoint pen into the foil wrapper along the tapered end of the chemlight, with the point of the pen barely extruding from the wrapper (the tight fit keeps both the pen and chemlight in place). This will put a spot of soft light on your notepad at the tip of your pen. The spot can be adjusted in size (from the size of a nickel to the size of a quarter or more) by moving the chemlight either forward or backward in the wrapper. One holds the entire contraption as if it were a fat pen when writing.

Another technique is to use a portable tape recorder. You can usually speak softly enough so you do not disturb the training. Time is required afterward, however, to document what is on the tape. Often this process requires considerable time. For best results, the transcription should be done as soon as possible after the field event.

To record time at night, use a watch with a large white face and big black hands and numerals. Quite often there is enough ambient light to read the time without resorting to a chemlight or turning on a flashlight under cover.

Other Night Research Tips

Do as soldiers recommend for night operations. Determine the best place to pack/carry each item of equipment and maintain this packing scheme every night. You cannot afford to search for your chemlight, pen, and gloves during field observations in total darkness. Use either a retractable ballpoint pen or pencil. Avoid using writing instruments that must be capped (e.g., a fountain pen or felt tipped pen), because the cap will invariably be lost or misplaced at night.

Bring night vision equipment so you can see what is happening at critical times. Night vision goggles, image intensification pocket scopes, and thermal sights work well for this purpose. If possible, obtain night photography equipment to provide a permanent visual record of training events. Photography helps you remember and document what happened and it also is an excellent means to show others what happened during training. Video records during daytime training serve the same purpose.

Make the observation plan simple, just as soldiers make their night plans simple. A complicated plan is likely to fail at night. Your ability to communicate effectively and easily with team members changes when darkness falls, and it becomes very difficult to make on-the-spot changes when things go awry.

Do not neglect the importance of insuring that you and your team members are warm and dry. Night observations are typically cold and damp events. Take the time to get the appropriate cold weather gear so your extremities stay dry and therefore warm. When your body is shivering or freezing, you can't execute research tasks well. One particular challenge is discovering the best glove or mitten that can be used with a pen at night. There may be no optimum solution for this.

Taking Notes in the Rain

Taking notes in wet weather is always a challenge. One method that works fairly well is to use a stenographic notepad (6 x 9 in.) that can be inserted into a water-resistant cover. The cover can be periodically opened for brief periods to take important notes (i.e., don't make unnecessary notes in the rain and jeopardize ruining crucial information). The choice of writing implements is critical, especially when the paper eventually becomes damp and soft. Using a medium-point ballpoint pen will usually allow you to get some ink on paper. Fine-point pens will tear into the paper, while the ink used in some felt-tipped pens will smear and run. Fortunately, training and experimentation is typically halted for a passing deluge. For most observers, the real challenge is taking notes in mist and light rain for extended periods, when there is no overhead cover nearby. A poncho is a highly adaptable piece of raingear that offers sufficient room and overhead cover for note taking. It is a very good idea to have one available, as it is to have an ample supply of notepads and pens. Never use a wet notepad for a second day. Instead, use a fresh one while letting the wet one dry thoroughly.

Protective Gear

Hearing protection and eye protection are definite requirements. Sometimes you will not be given access to the training site without them. Safety goggles with side impact protection and prescriptive lenses, if needed, are indispensable. Either foam or fitted earplugs are inexpensive and effective means of hearing protection. Helmets are required whenever live firing is being conducted. In addition to several eye injuries, heat exhaustion was the most common training casualty suffered by MOUT ACTD participants. Thus, don't forget to bring sunglasses, sunscreen, a hat, and plenty of fluids whenever going to the field. It is not always possible to predict how long you will be there.

Field Office Essentials

During periods of extended field observation, one's personal vehicle will tend to become one's office, or at least one's supply room. The following list contains recommendations for supplying that office:

Snack food and beverages (your next opportunity for a meal may not be predictable)
Ample supply of notepads (stenopads work well), pens, chemlights, and other consumables
Night vision goggles and thermal observation devices (if available), including hand receipts
Notebook computer and peripherals
Raingear and cold weather clothing (light jacket, heavy coat, and insulated gloves at a minimum)
Boots (both summer and winter varieties)
Briefcase to carry important files and documents (for what you bring to and take from the field)
Cellular phone
Multi-function tool or small toolbox
Back-up method of time keeping (especially if your watch battery isn't new)
Spare batteries for all electronics
Dry towels (paper or cloth)
Disposable wet napkins are good for improvised cleaning
Insect repellant (the stronger the better)
Cash and credit cards, to purchase what you forgot to bring
Protective gear listed in previous section
Small portable office kit (e.g., a plastic box containing common office tools such as pencils, pens, ruler, paper clips, cellophane tape, 3x5 cards, scissors, rubber bands, stapler, staples, self-stick removable notes, felt-tip markers)
Camera (photographs help you to better document events, to more easily recall those events, and to more effectively describe them to others)
Camcorder (for recording action events)

Field Research Etiquette

One's behavior and conduct in the field may be interpreted differently than it is in institutional settings. The following recommendations will help to insure your presence in the field will be welcomed by others.

Arrive early and be prepared (at least act like you are if you aren't).
Don't rely on research participants for subsistence items (instead, bring a surplus and share).
Make a comprehensive list of your research requirements and present them to those in charge beforehand (you won't make friends thinking of new ones as you go along).
If you have a question about what was said or why something was done, ask it later during a break (don't interrupt the training, as you should be as unobtrusive as possible).
Remain on-site for the entire duration of training or experimentation.
Adopt the attitude of an impartial observer (you're there to learn, not to evaluate).
When asked what you are doing (we guarantee you will eventually be asked), explain that you are looking at how different technologies and systems impact training. Also make it clear that you are evaluating the technologies themselves, and are not evaluating the performance of individuals or groups. Most soldiers and instructors will freely share important information that you may have otherwise missed, once they better understand what you are doing and why you are doing it.

Appendix B

Observation Form

DATE: ________ TRAINING DAY [1st - nth] DATA COLLECTOR __________

WEATHER CONDITIONS [*Temperature, precipitation, humidity, etc.*]

TASK/PERIOD [*Task, skill, content, and/or activity; Instruction or Test*].

MODE OF INSTRUCTION

[*Lecture, demonstration, hands-on practice, question-answer, one-on-one, ...]*

INSTRUCTOR TO STUDENT RATIO [1 to xx]

TOTAL INSTRUCTIONAL TIME [xx min]

START TIME [:] STOP TIME [: [

TIMES FOR SPECIFIC TOPICS [*Summary of times cited under Instructional Record below. Includes breaks in training. Times sum to Total Instructional Time above.*]

[Topic or Area #1]	START [:]	STOP [:]
[Topic or Area #2]	START [:]	STOP [:]
[Topic or Area #3]	START [:]	STOP [:]
[Topic or Area #n]	START [:]	STOP [:]

INSTRUCTIONAL/TRAINING RECORD

[*Observer's record of instruction/training. Recorded in sequence by critical time segments; breaks recorded with reason for break cited. When lecture format, content presented was recorded. Details of instructor demonstrations recorded. ....*]

TRAINING DEVICES

[List of training devices; quantities of each device.]

TRAINING AIDS

[*List of training aids; e.g., Army films/tapes, student handouts, visual aids of various types, models of threat and enemy vehicles, overhead projection equipment, actual equipment, ...*]

STUDENT PERFORMANCE DATA [*Time and error data. Recorded during hands-on practice or testing.*]

			# Errors	
[Student name/#]	ST[__:__]	STP[__:__]	________	Go/NoGo
[Student name/#]	ST[__:__]	STP[__:__]	________	Go/NoGo
[Student name/#]	ST[__:__]	STP[__:__]	________	Go/NoGo
[Student name/#]	ST[__:__]	STP[__:__]	________	Go/NoGo

TRAINING PROBLEMS

[*Record of incidents that created training problems, e.g., equipment failures, ammunition not on time, poor lighting, poor quality videotapes, failure to execute procedures correctly, ...]*

PREREQUISITE KNOWLEDGE AND SKILLS ASSUMED

[*Skills and knowledge students need to understand material presented or to perform the task.]*

QUALITY OF INSTRUCTION

[*Comments on training from an instructional design perspective; problems encountered by students in mastering materials; etc.*]

OTHER NOTES/OBSERVATIONS

FOR TESTS ONLY

Proficiency Test administered immediately after instruction? Yes No

Proficiency Test: Written Hands-on None Other: __________

Proficiency Standards: Go/NoGo Numeric None Other:

[*Additional information on tests such as test items, quality of instructor testing, equipment and aids used. Number of attempts allowed was also recorded.*]

TEST PROCEDURES

TEST [*Name of task or skill tested*]

DESCRIPTION OF TEST PROCEDURE

[*Details on test instructions, equipment used, sequence of events, type of test (written, hands-on, etc.), whether test was group or individually administered, ...*]

CRITERIA FOR PASSING

DID TEST CORRESPOND TO TRAINING OBJECTIVE: Yes No
If NO, describe discrepancies.

WERE TEST DIRECTIONS CLEAR, SIMPLE, AND EASY TO FOLLOW? Yes No

WAS THE TEST DIFFERENT FROM THE PRACTICE AND/OR EXAMPLES GIVEN DURING THE CLASS? Yes No
If YES, how did the test differ?

RESULTS OF TESTING

[List of test results by students - test time, score, Go/NoGo, # trials]

SUMMARY OF RESULTS

Time allocated per student ________
Total testing time for class ________
Average score _____

Min/max times for students _____/_____
Average time/student _____
Average # errors/student _____

Types of errors [*List of types of errors and their frequency*]

Number of students who passed first time _____
Number of students of failed first time_____

Number of students who finally passed _____
Number of students who passed but required retests _____
Number of students who failed _____